**Technical Write-Up: Wildfire Analysis and Response Efficiency**

**1. Title & Introduction**

* **Project Title**: Analysis of Wildfire Trends, Causes, and Response Efficiency
* **Introduction**: A comprehensive examination of wildfires, their growing impact, leading causes, and response efficiency, drawing on extensive data from the US Forest Service.

**2. Problem Statement**

* The increasing frequency and severity of wildfires necessitates a thorough understanding of their trends, causes, and the effectiveness of the response measures in place.

**3. Data Source & Collection**

* **Data Source**: USFS (United States Forest Service)
* **Description**: The dataset comprises over 2 million rows of individual wildfires, detailing fire size (from near-zero to over 600,000 acres), date, cause, location, and more.
* [Forest Service Research Data Archive (usda.gov)](https://www.fs.usda.gov/rds/archive/catalog/RDS-2013-0009.6)

**4. Data Cleaning & Pre-processing**

* Encountered missing data; however, most of the analysis was conducted using the available data. Further details on data cleaning and exploration in submitted documentation.

**5. Exploratory Data Analysis (EDA)**

* In python I used df.describe() , df.isnull() and df.dtypes to get an overview of the dataset.
  + Found significant amounts of nulls. However, these weren’t all necessary for my analysis of other fields. Also, when basing my analysis mostly on Class G fires, I found that the nulls were only apparent for insignificant fire classes that were not included in my analysis.
    - A computer screen shot of a black screen

      Description automatically generated
  + Identified that Fire\_Size’s MAX size is considerably larger than the 75th percentile, indicating some very large fires skew the averages which helped me understand the need to focus on categorizing the fires.
    - A black screen with white text

      Description automatically generated
  + Initial exploration showed that the amount of fires were greatly skewed toward small fires while the larger fires were responsible for more size, which would need to be accounted for in analyses.
    - A screen shot of a graph

      Description automatically generated

**6. Methodology & Models**

* Analysis mainly focused on Class G fires for their significant contribution to total acreage burned.
* Used Tableau dashboards to visualize causes, locations, and durations of these wildfires.
* Introduced the Control Efficiency Score (CES) to measure the efficiency of landowners in controlling fires.

**Control Efficiency Score (CES): Definition and Importance**

* The Control Efficiency Score (CES) was formulated to provide a metric to measure the efficiency of landowners in controlling wildfires. With the growing concern surrounding the frequency, size, and impact of wildfires, there's a crucial need to assess how efficiently these fires are managed and controlled. CES offers a quantifiable metric to compare different landowners' performances and gain insights into potential areas for improvement.

**Formulation of CES**:

* CES evaluates the average acres controlled per hour by landowners. The idea behind this formulation was to account for both the size and duration of the fire, thus providing a comprehensive measure of efficiency. The formula for CES is: CES=Total Acres Burned / Total Duration (in hours)

**Findings from the Analysis**:

* Based on the CES, various landowners showcased varying efficiencies. For instance:
  + The Bureau of Land Management (BLM) showcased a higher efficiency in controlling fires compared to the United States Forest Service (USFS), when accounting for the number of class G fires they managed.
* However, looking at the acres burned per hour, BLM had a higher rate, indicating that while they controlled larger fires more efficiently, the fires under their watch grew at a faster rate.

**Limitations and Considerations**:

* While CES is a valuable metric, it's important to consider its limitations:
* External Factors: CES does not account for external factors that might affect fire control, such as weather conditions, terrain difficulties, availability of firefighting resources, and others.
* Depth of Data: The CES metric is based on the data provided, which might not capture all relevant details about each fire.
* Uniformity Assumption: CES assumes equal importance for all fires, not considering the potential harm, location, or other contextual factors which might make one fire more critical than another.

**Potential Applications**:

* CES can be a foundational tool for policymakers, landowners, and firefighting agencies. By identifying high-performing and low-performing landowners, resources and training can be allocated more effectively. Furthermore, CES can be a starting point for more in-depth investigations into why certain landowners perform better or worse than others.

**7. Key Findings**

* Despite the majority of fires being man-made, naturally caused fires account for a higher total burned area.
* Geographic analysis revealed pronounced wildfire activity in the West and South, with notable intensities in states like California and Florida.
* Identified a rising trend in the number and size of wildfires over the years.
* A graph of red lines

  Description automatically generated with medium confidence
* Found that while the majority of fires are human-caused, natural causes account for 58% of the total acreage burned. Arson is particularly prevalent for acres burned after Natural and Missing causes.
* A screenshot of a computer

  Description automatically generated
* Ownership visualization with calculated efficiency score:
* A screen shot of a data

  Description automatically generated
* Created geographic visualizations that showed fire seasonality through regions of the US, displaying clear patterns.
* A screenshot of a computer

  Description automatically generated
* Created geographic visualizations that showed cause seasonality as well.
* A screenshot of a graph

  Description automatically generated

**8. Recommendations & Next Steps**

* Landowners should continually strive to improve efficiency in combating wildfires.
* Resources should be allocated seasonally, based on predictable wildfire patterns.
* Public awareness campaigns can significantly reduce the human-induced wildfire percentage.
* Further studies could incorporate temperature data for predictive models, expand geographic scope, and work with high CES landowners for enhanced training.

**9. Limitations & Assumptions**

* The dataset may not be fully comprehensive, and some incidents could be missing or unreported.
* Data integrity could be compromised by unspecified or missing labels.
* The Control Efficiency Score, while informative, doesn't account for all variables impacting firefighting efficiency.

**10. Conclusion**

* The rising trend in wildfires underscores the urgency to bolster response measures and public awareness. The data-driven insights presented herein serve as a foundation for actionable recommendations and strategies to mitigate the wildfire crisis.

**11. Code & Tools**

* **Tools & Libraries**: Tableau Public for visualization, Python (Pandas, SQLite, Seaborn) for EDA
* **Code Access**: The complete set of dashboards and analysis can be accessed on [Tableau Public](https://public.tableau.com/views/WildfireAnalysisCapstoneGeneralAssembly/Story1?:language=en-US&:display_count=n&:origin=viz_share_link) and the documentation on [GitHub](https://github.com/drewcollie).